

5. Dameison, G. J. and Moore, G. T., *Bull. Am. Assn. Petrol. Geol.*, 1980, **64**, 1179.
 6. Hallouty, M. T., *Bull. Am. Assoc. Petrol. Geol.*, 1980, **64**, 1210.
 7. *Sci. American*, 1980, **242**, 130.
 8. *Nature (London)*, 1982, **300**, 312.
 9. Wakita, H. and Sano, Y., *Nature (London)*, 1983, **305**, 793.

ON POSITIVE AND NEGATIVE VALUED AUXILIARIES IN SURVEY SAMPLING

T. SRIVENKATARAMANA and N. G. N. PRASAD*

Department of Statistics, Bangalore University, Bangalore 560 056, India.

**Carleton University, Ottawa, Ontario, Canada.*

SUMMARY

Auxiliary variates in sample surveys are occasionally positive and negative-valued. This introduces difficulties in using ratio or product methods of estimation. A simple translation is suggested to overcome the problem. Simple random sampling is assumed for illustration.

1. INTRODUCTION

CONSIDER estimating the total Y of a character y in a finite population of N units, based on a probability sample of n ($\leq N$) units. For illustration assume simple random sampling without replacement and let \bar{y} be the sample mean. The commonly used ratio and product estimators are

$$\hat{Y}_r = N\bar{y}(\bar{X}/\bar{x}), \quad \hat{Y}_p = N\bar{y}(\bar{x}/\bar{X}); \quad (1.1)$$

where \bar{x} is the sample mean of a character x auxiliary to y and it is assumed that the population mean \bar{X} is known. Generally x is non-negative. But exceptions do occur. Examples are variates arising as differences, like savings and net revenue, which can be positive or negative. These variates cannot be directly used as auxiliaries in (1.1) since \bar{X} and/or \bar{x} may be close to zero.

The present note considers a simple translation of x to make it positive-valued, along with a shift for y so as to have proportionality between the values of the two variates. Thus define

$$u = x + a, v = y + b; \quad (1.2)$$

where a is chosen such that $u > 0$. For instance, let x_1 (< 0) be the smallest x -value. Then a must satisfy $|x_1| < a < \infty$. Thus the magnitude of the smallest x -observation in the population must be known at least approximately in order to fix the choice for a . This should pose no problem since x is an auxiliary character on which information is supposed to be available easily. Next, b is chosen to control the mean

squared error (MSE) of the estimator, as discussed in Section 3.

2. THE ESTIMATOR

Let $\bar{u} = \bar{x} + a$, $\bar{v} = \bar{y} + b$, $\bar{U} = \bar{X} + a$, $\bar{V} = \bar{Y} + b$. Then the usual ratio estimator of the total Y is

$$\hat{V}_r = N\bar{v}(\bar{U}/\bar{u}) = N(\bar{y} + b)(\bar{X} + a)/(\bar{x} + a) \quad (2.1)$$

and hence Y may be estimated by

$$\hat{Y}_1 = \hat{V}_r - Nb. \quad (2.2)$$

The transformations (1.2), being only changes of origin, leave the variances, covariances and correlations unchanged. Also the standard theory for ratio estimators applies to \hat{Y}_1 . Thus the bias and MSE of this estimator are, up to second order moments,

$$B(\hat{Y}_1) = (N - n)(RS_x^2 - S_{xy})/n(\bar{X} + a), \quad (2.3)$$

$$M(\hat{Y}_1) = N(N - n)(S_y^2 + R^2S_x^2 - 2RS_{yx})/n, \quad (2.4)$$

where $R = (\bar{Y} + b)/(\bar{X} + a)$ and S_y^2 is the population variance of y , etc.

For given a , the b minimizing the MSE of the estimator \hat{Y}_1 upto second order moments can be obtained by differentiating with respect to b the expression for the MSE in (2.4) and equating it to zero. This leads to the condition $R = S_{yx}/S_x^2$ which implies $b_{opt} = B(\bar{X} + a) - \bar{Y}$ where $B = S_{yx}/S_x^2$ is the coefficient of regression of y on x in the population. For this choice of b , \hat{Y}_1 is almost unbiased for Y . And

$$M_{min}(\hat{Y}_1) = N(N - n)S_y^2(1 - \rho^2)/n,$$

where ρ is the population correlation coefficient. This MSE is the same as the large sample approximation to the MSE of the linear regression estimator. A geometric interpretation of b_{opt} is in figure 1. In particular, for the special case $y_i = Bx_i$ for all i , $b_{opt} = Ba$ and \hat{Y}_r reduces to $NB(\bar{X} + a)$ so that $\hat{Y}_1 = NB\bar{X}$. Hence \hat{Y}_1 estimates $Y = NB\bar{X}$ without any error, just like \hat{Y}_r in this situation.

However, in survey practice, the optimum choice for b can rarely be implemented exactly since this requires the knowledge of the population parameters B and \bar{Y} . Therefore as an alternative we find the range of values for b which keep \hat{Y}_1 more precise than the simple estimator $N\bar{y}$. In (2.4) the first term corresponds to the sampling variance of $N\bar{y}$. Hence \hat{Y}_1 will be more precise than $N\bar{y}$ as long as $R^2 S_x^2 - 2RS_{yx} < 0$. That is $R(R - 2B) < 0$. This reduces to $B/R > 1/2$ or, as a condition on b , that

$$b \text{ lies between } -\bar{Y} \text{ and } D - \bar{Y} \quad (2.5)$$

where $D = 2B(\bar{X} + a)$. This gives the required range of values for b . Further, the introduction of the parameter b will be really justified only if it makes the estimator more precise than that without b (i.e. with $b = 0$). This consideration requires that $R^2 S_x^2 - 2RS_{yx} < R_0^2 S_x^2 - 2R_0 S_{yx}$ where $R_0 = \bar{Y}/(\bar{X} + a)$, leading to the condition that

$$b \text{ should lie between } 0 \text{ and } D - 2\bar{Y}. \quad (2.6)$$

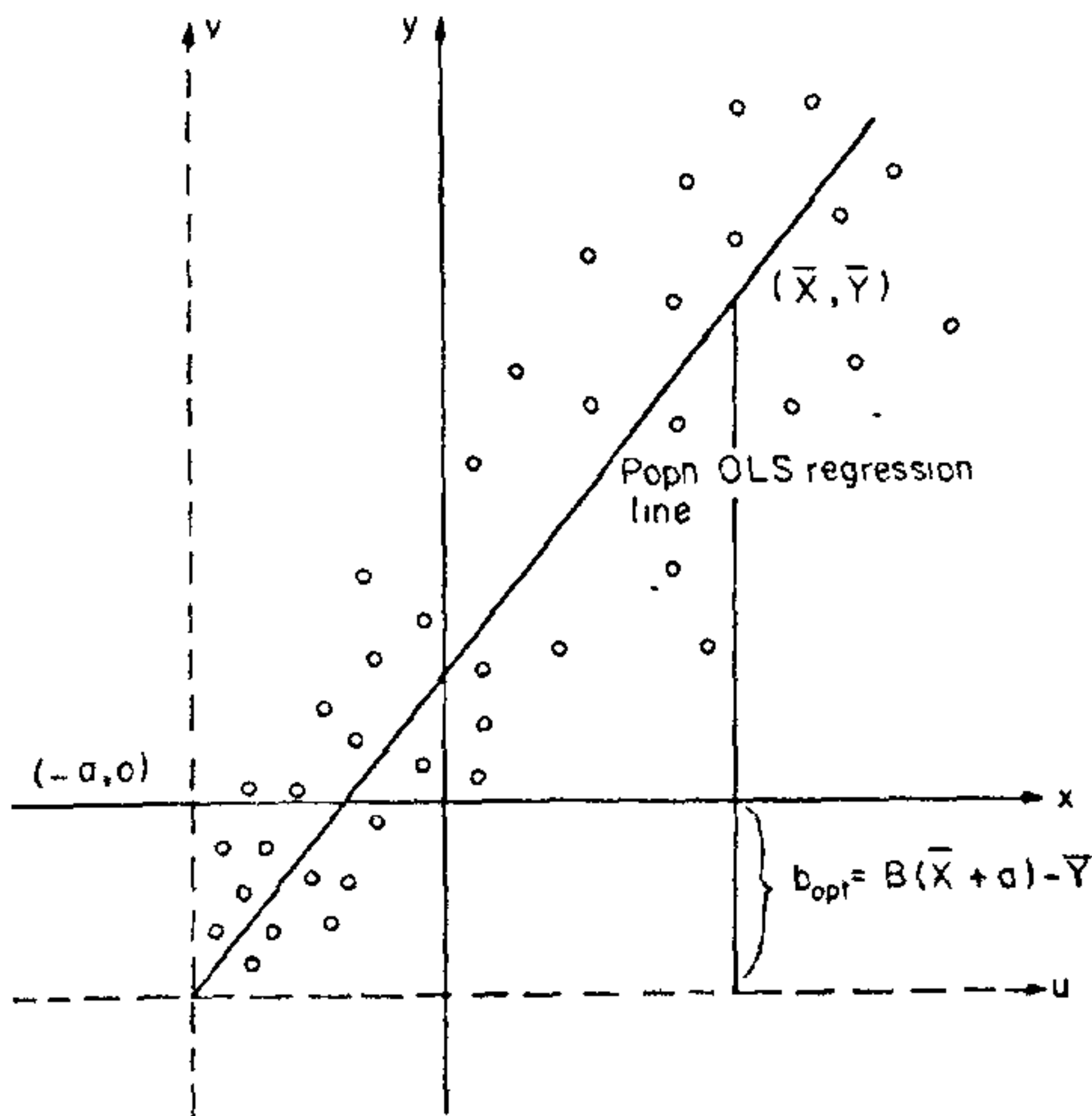


Figure 1. Geometric interpretation of b_{opt} .

This condition also ensures a simultaneous reduction in the absolute bias relative to the $b = 0$ case. Thus (2.5) and (2.6) specify two intervals inside which b should lie in order that \hat{Y}_1 is more precise than both $N\bar{y}$ and \hat{Y}_r with $b = 0$. The common midpoint of these intervals is b_{opt} , while the widths are $|D|$ and $|D - 2\bar{Y}|$ respectively.

3. THE CHOICE OF b IN PRACTICE

It was pointed out that the ideal choice for b is b_{opt} which is also the midpoint of the intervals specified in (2.5) and (2.6). However hitting precisely this value is difficult since it depends on the unknown parameters B and \bar{Y} . In this context past experience and data from a pilot study may be helpful to find approximations to B and \bar{Y} and hence to b_{opt} . Alternatively, since b is needed only at the estimation stage of a survey, the knowledge of the sample mean \bar{y} and a scatter diagram for a part of the sample data on y and x may provide some guidance. Let us also examine the sensitivity of $M(\hat{Y}_1)$ to deviations of b from its optimum value. Suppose it is stipulated that the relative increase in the MSE as compared to $M_{min}(\hat{Y}_1)$ should not exceed c . That is

$$[M(\hat{Y}_1) - M_{min}(\hat{Y}_1)] / M_{min}(\hat{Y}_1) \leq c. \quad (3.1)$$

Writing $R = B(1 + d)$, it can be shown that (3.1) reduces to

$$d^2 \leq c(1 - \rho^2) / \rho^2. \quad (3.2)$$

Thus for small values of ρ^2 , $d = (R - B)/B$ can be considerable in absolute value and yet satisfy (3.2). For these values of ρ^2 , R can depart considerably from B , and hence b from b_{opt} , without resulting in a substantial increase in the MSE compared to the minimum. But this flexibility disappears as ρ^2 approaches unity.

4. THE PRODUCT METHOD

When the character of interest, y , is negatively correlated with the auxiliary character x we may consider the product estimator^{1,2}. This in fact admits closed form expressions for bias and MSE, unlike the ratio method. After the translations of y and x suggested above the product estimator will be of the form

$$\hat{Y}_2 = N(\bar{y} + b)[(\bar{x} + a)/(\bar{X} + a)] - Nb. \quad (4.1)$$

This has

$$B(\hat{Y}_2) = (N - n)S_{yx} / n(\bar{X} + a), \quad (4.2)$$

$$M(\hat{Y}_2) = N(N - n)[S_y^2 + R^2 S_x^2 + 2RS_{yx}] / n. \quad (4.3)$$

In this case $b_{\text{opt}} = -B(\bar{X} + a) - \bar{Y}$. The two conditions on b , corresponding to (2.5) and (2.6), are that it lies between $-\bar{Y}$ and $-(D + \bar{Y})$, and 0 and $-(D + 2\bar{Y})$. As compared to the usual product estimator $\hat{Y}_p = N\bar{y}(\bar{x}/\bar{X})$, there will be a reduction in the absolute bias if and only if $|\bar{X} + a| > |\bar{X}|$. This will always hold when $\bar{X} \geq 0$ since $a > 0$. If $\bar{X} < 0$, it becomes necessary that $a > 2|\bar{X}|$.

ACKNOWLEDGEMENT

The authors wish to thank the referee for his very helpful comments on the first version of this paper.

21 February 1983; Revised 12 January 1984

1. Robson, D. S., *J. Am. Stat. Assoc.*, 1957, 52, 511.
2. Murthy, M. N., *Sankhya*, 1964, A26, 69.

POLYCYCLIC AROMATIC COMPOUNDS: SYNTHESIS OF POLYCYCLICS CONTAINING ANTHRACENE, PHENANTHRENE AND FLUORENE NUCLEUS

S. MONDAL and A. J. BHATTACHARYA

Department of Chemistry, University of Burdwan, Burdwan 713 104, India.

ABSTRACT

A new series of polycyclic aromatic compounds with anthracene, phenanthrene and fluorene nucleus have been prepared. The method involves the Diel's-Alder cycloaddition of cyclopentadienones (I, II, III) with the dienophiles, e.g. 1-chloro-3,4-dihydronaphthylene (VI) and 5,6-dimethyl-1-chloro-3(H)-indene (V). The dienophiles V and VI are prepared from the corresponding cyclic ketone with phosphorus pentachloride.

INTRODUCTION

CARCINOGENIC activity found in certain polynuclear hydrocarbons with a certain degree of molecular complexity and chemical reactivity prompted us to devise a synthetic method for a new series of highly substituted aromatic hydrocarbons. 1:2, 3:4, 5:6-tribenzanthracene¹ and 1,2,3,4-tetramethylphenanthrene are feebly active. Replacing two methyl groups in the latter by a benzene ring leads to more active 1,2-dimethylchrysene. It is observed that substitution both on the phenanthrene and on anthracene nucleus plays an important role on their physiological activities.

In the present work, the chloro olefinic compounds, 1-chloro-3,4-dihydronaphthylene (VI) and 5,6-dimethyl-1-chloro-3(H)-indene (V) were synthesised and successfully used as dienophiles in the diene synthesis. A new series of 1,2,3,4-tetraarylphenanthrene and 9,10-diaryl-1:2, 3:4, 5:6-tribenzanthracene derivatives were synthesised in order to evaluate the effect of various aryl groups in the 1,2,3,4-position of phenanthrene nucleus and in the 9,10-position of tribenzanthracene nucleus.

MATERIALS AND METHODS

The method involves Diels-Alder cycloaddition reaction of various cyclopentadienones, viz 1,3-diaryl-2(H)-cyclopenta[1]-phenanthrene-2-one [Ia, Ar = Ph; Ib, Ar = *p*-CH₃C₆H₄; Ic, Ar = *p*-C₂H₅C₆H₄; Id, Ar = 3,4-(CH₃)₂C₆H₃], 2,5-diaryl-8(H)-oxo-cyclopent[a]-acenaphthylene (IIa, Ar = Ph; IIb, Ar = *p*-OCH₃C₆H₄) and 1,4-diaryl-2,3-diphenyl-cyclopentadienone (IIIa, Ar = Ph, IIIb, Ar = *p*-OCH₃C₆H₄) with 1-chloro-3,4-dihydronaphthylene (VI). It is believed that the reaction of the chloro compounds (V, VI) with cyclopentadienones follows the Alder-Endo rule². The cycloaddition reaction of cyclopentadienones with V and VI could follow two courses. However, in general, only one of these is actually realised. After preorientation of the reactants, the dienophile is added in such a way as to give the maximum concentration of π -bonds in the transition state. Maleic anhydride forms exclusively an *endo*-adduct with cyclopentadiene³. The formation of an *endo*-adduct is also found in the condensation of cyclopentadienones with *p*-benzoquinone⁴. The dienophile (VI) was conveniently prepared by the